AD ______________

REPORT NO T98-9

EVALUATING RISK OF REINJURY AMONG 1,274 ELITE ARMY AIRBORNE SOLDIERS

U S ARMY RESEARCH INSTITUTE OF ENVIRONMENTAL MEDICINE
Natick, Massachusetts

JANUARY 1998

Approved for public release: distribution unlimited

UNITED STATES ARMY
MEDICAL RESEARCH AND MATERIEL COMMAND

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED
Technical Report

4. TITLE AND SUBTITLE
Evaluating risk of reinjury among 1274 elite Army Airborne Soldiers

5. FUNDING NUMBERS

6. AUTHOR(S)
G.A. Schneider, P. Amoroso, and C. Bigelow

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute of Environmental Medicine
Natick, MA 01760-5007

8. PERFORMING ORGANIZATION REPORT NUMBER
T98-9

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, MD 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Adapted from a Masters Thesis University of Massachusetts - Amherst School of Public Health and Health Sciences

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)
Risk of reinjury is poorly quantified in the medical literature. Two general factors may compete to alter an individual's risk of reinjury. While an injured individual may be unable to participate in normal activities thereby lowering exposure to injurious activities, an injured individual may also be more likely to be injured if insufficient recovery time is possible after the initial injury. The objective of this report was to model the risk factors for a second injury given that a soldier had already been injured once. The study used a retrospective cohort design to follow 1214 airborne soldiers for periods up to 18 months. Records were collected from multiple sources including, battalion aid station outpatient records, dental records, jump logs, unit physical fitness test scores, unit height and weight rosters, and unit personnel records. Several time to event statistical techniques were evaluated but only one is reported here. Because both the baseline hazard and the parameter estimates are allowed to vary by injury event, the best technique was found to be the Prentice, Williams, and Peterson (PWP) variation of the Cox proportional hazards model. Risk of injury was significantly higher among individuals who had already been injured once. This reinjury risk was highest among individuals of Hispanic race, users of alcohol, married soldiers, individuals seen only by a medic for the first injury, and among younger and less physically fit soldiers.

14. SUBJECT TERMS
Airborne, injuries, Cox proportional hazards model, tobacco, alcohol, reinjury, parachuting, infantry, race, marital status

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Risk of Reinjury Among 1,214 Elite Army Airborne Soldiers

Prepared by

Gary A. Schneider, MS
Paul J. Amoroso, MD, MPH

U.S. Army Research Institute of Environmental Medicine

Carol Bigelow, PhD

University of Massachusetts, Amherst

Military Performance Division
U.S. Army Research Institute of Environmental Medicine
Natick, MA 01750-5007

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

BACKGROUND

This report provides partial and preliminary results from a larger study of the health of a brigade of Airborne Soldiers at Fort Bragg, North Carolina. The content of this report was distilled from a thesis which also served as partial fulfillment of the degree requirements for the Masters of Science degree in Biostatistics at the University of Massachusetts, Amherst. The full thesis was also published as ARIEM Technical Note 98-1, entitled "Multiple Event Analysis of Injuries Using Adaptations to the Cox Proportional Hazards Model."

ACKNOWLEDGMENTS

We wish to thank the following people:

Dr. Penelope Pekow and Dr. David Hosmer from the biostatistics department in the School of Public Health and Health Sciences, University of Massachusetts, Amherst, for their aid in the completion of this project.

Anthony Pusateri, Joseph Dettori, Bruce Jones, Katy Reynolds, Everett Harmon, Michael Smutok, and John Patton from the Army Research Institute of Environmental Medicine, for their constant support, as well as Michelle Yore, Frank Hoin and Shari Hallas for their various contributions toward the completion of this project.

The student interns and enlisted soldiers who aided in the painstaking tasks of the collection and entry of the data used for this project.

The highly trained and dedicated soldiers of the 82nd Airborne Division who continuously stand ready to risk their lives in the service of their country.

EXECUTIVE SUMMARY

Risk of reinjury is poorly quantified in the medical literature. Two general factors may compete to alter an individual's risk of reinjury. While an injured individual may be unable to participate in normal activities, thereby lowering exposure to injurious activities, an injured individual may also be more likely to be injured if insufficient recovery time is possible after the initial injury. The objective of this report was to model the risk factors for a second injury given that a soldier had already been injured once. The study used a retrospective cohort design to follow 1214 Army Airborne soldiers for periods up to 18 months. Records were collected from multiple sources including battalion aid station outpatient records, dental records, jump logs, unit physical fitness test scores, unit height and weight rosters, and unit personnel records. Several time to event statistical techniques were evaluated, but only one is reported here. Because both the baseline hazard and the parameter estimates are allowed to vary by injury event, the best technique was found to be the Prentice, Williams, and Peterson (PWP) variation of the Cox Proportional Hazards Model. Risk of injury was significantly higher among individuals who had already been injured once. This reinjury risk was highest among individuals of Hispanic race, users of alcohol, married soldiers, individuals seen only by a medic for the first injury, and among younger and less physically fit soldiers.

INTRODUCTION

Individuals in the military, by virtue of their occupation and demographic profile, are at high risk for injury. The reduction in productivity and economic effects of injury are profound; however, there is still an incomplete understanding of the many risk factors for injury. A recent effort by the Armed Forces Epidemiology Board culminated in the publication of Injuries in the Military - A Hidden Epidemic. This report provides a comprehensive review of the extent of the injury problem for the United States Military (Jones and Hanson, 1996). Within this report it is suggested that "previous injury history" and the "late effects of injury" have an effect on subsequent risk of injury. These conclusions resulted from review of hospitalization data and results from epidemiologic risk factor studies in which previous injury history was determined by self reporting. Additionally, the two cited epidemiological studies in which previous injury history was examined have contradictory findings. One suggests that previous injury history is a risk factor for later injury (Jones, Cowan, Tomilson, et al., 1993). The other suggests a protective effect (Brodine and Shaffer, 1995). Such disagreement suggests further inquiry concerning the effect of previous injury on subsequent injury is needed. Increasing the understanding of the effects of previous injury history is therefore one of the primary objectives of this report.

The most comprehensive way to examine previous injury history as a potential risk factor for injury would be to prospectively follow a cohort of individuals for a time interval of adequate duration so that there would be a subgroup of subjects who experienced two or more injuries. Comprehensive collection of all injury data should be extracted from the individual's medical record in order to ensure maximum ascertainment of injury events. There have been no published risk factor studies conducted on military populations in which previous injury history was determined from the individual's medical records.

Prospective examination of the population would provide many benefits. Both type and severity of previous injury could be evaluated, as well as multiple aspects of the medical care provided. These possible risk factors can be examined concurrently with previously identified risk factors. Additionally, the effect of elapsed time since previous injury on subsequent injury can be explored.

The statistical tools needed for the proper analysis of recurrent injury data require sophisticated techniques that are largely untested for this purpose. This is primarily because the addition of the medical record review to assess previous injury history necessitates the study of a dynamic cohort of individuals, some of which may not have an injury during the study interval. This analysis is further complicated by the multiple injury setting. Regression techniques that have been previously used to explore the multiple injury setting are the Andersen-Gill Model, the Prentice, Williams, and Petersen (PWP) Model, and a Cox Proportional Hazards Model for an individual's final injury. Previous research has demonstrated that the combination of the latter two models are needed to comprehensively examine this situation (Schneider, In Press).

Therefore, an additional focus of this report is to provide the reader with an understanding of these regression techniques as they pertain to multiple injury data analysis. Detailed explanations of these statistical models can be found in Appendices 1 and 2.

METHODS

DATA COLLECTION

The objective of this study was to collect retrospective data from a variety of sources on a dynamic population of Army Airborne soldiers in order to conduct a comprehensive morbidity evaluation. In order to accomplish this a relational database was designed and tested using EpiInfo version 6.0 prior to data collection. A "parent" file was constructed during October 1994 using an electronic personnel roster from one brigade of the 82nd Airborne Division (n=2147), Fort Bragg, NC. The roster was obtained from Division Headquarters. A four digit unique identifier was created for each individual.

Collection and abstraction of study data occurred during seven visits to Fort Bragg, NC, between November 1994 and March 1996. Several categories of data were collected, each maintained in a different location. Information in the Annual Health Questionnaires for Dental Treatment was located in the dental clinic; however, each battalion had its own medical clinic where individual medical records were located. Army Physical Fitness Test (APFT) score cards were housed in the company area, of which there were five per battalion (15 in total).

Data were abstracted by making photocopies of each individual's record from each data source. Data from these photocopied records were then entered into EpiInfo and linked electronically to the parent file via the four digit unique identifier. Data sources were each individual's outpatient medical records, Annual Health Questionnaire for Dental Treatment, APFT score card, and individual jump logs (not discussed or analyzed for this report). Additionally, demographic data were extracted from the Total Army Injury and Health Outcomes Database (TAIHOD). Details on the data collection for each data source are given below.

Parent File

The parent file was constructed during October 1994 by obtaining an electronic roster of one brigade in the 82nd Airborne Division (n=2147). This brigade consisted of three, 671-person battalions and a 134-person headquarters company. The original intent of this research was to conduct a comprehensive morbidity evaluation of the entire brigade; however, logistical, budgetary and personnel constraints allowed the medical record reviews to be conducted on only two battalions (n=1342). Therefore, for the purposes of the data presented in this report, the available size of the parent file was limited to the individuals in these two battalions.

The dynamic nature of this population made it necessary to make some changes in the target population. Ninety-four (n=94) subjects were added to the parent file because the presence of records on these persons from one or more of the data sources indicated they were missing from the original parent file. Likewise, there were 162 subjects for whom data were unavailable from *any* data source. These subjects were considered "non-arrivals" and were thus deleted from the parent file. This resulted in a net reduction of the functional size of the parent file (1342+94-162=1274).

For the calculation of survival times, a roster of each individual's arrival date to the brigade was constructed by the Brigade Headquarters during March 1995. This enabled the calculation of each subject's time contribution to the study as the number of days between an individual's arrival date and January 31, 1995, the last day of the study interval. Each individual's time contribution was limited by the length of the medical record review which, was 396 days (13 months) for first battalion and 549 days (18 months) for second battalion. If an individual's arrival date occurred prior to the beginning date of the medical record review, their person time was truncated to the maximum allotted for their respective battalion. The arrival date to the brigade was available on all but 60 (4.7%) of the 1274 individuals on the parent file. Thus, the final analysis sample size for these analyses is n=1214.

Medical Records

The original intent of a February 1995 data collection trip was to conduct outpatient medical record reviews on the entire brigade for the 18- month interval, ending on January 31, 1995. However, logistical, budgetary and personnel constraints limited the review of the medical records to 2 of the 3 battalions. Additionally, the medical record review of one of these battalions was intentionally limited to 13 months.

Information specific to a medical problem was recorded on a pre-designed data collection form that included diagnosis; body part (if an injury); physiological system (if an illness); number of follow-up visits; and highest level of medical provider seen for the problem. Efforts to record lost duty time were abandoned due to the inconsistent and incomplete nature of the documentation of limited duty profiles within the medical records.

Nine hundred-eighty (n=980, 80.7%) individual medical records were reviewed of the 1214 subjects remaining in the functional parent file, during February 1995. An additional 185 (15.2%) of the medical records were reviewed during one of four data collection trips that occurred between March and July of 1995, yielding a total of 1165 (96.0%) medical records which received complete review.

Dental Records

Within each individual's dental record is a Health Questionnaire for Dental Treatment that is updated at least annually at the time of an individual's routine dental checkup or visit for acute care. This questionnaire consists of 33 questions in which the individual can answer yes, no, or unknown (Appendix 3). Two of these questions pertain to cigarette use and alcohol use and were the primary reason for abstracting these data (Amoroso, 1996).

One thousand five hundred ninety (n=1,590) dental questionnaires, representing approximately 74.0% of the brigade, were initially photocopied in November 1994. An additional 398 (18.5% of the brigade) dental questionnaires were photocopied during one of four additional data collection trips between February and July of 1995, yielding a total of 1988 collected dental questionnaires. Of the 1214 subjects in the functional parent file, dental questionnaire data were collected on 1163 (95.8%).

Physical Fitness Data

The APFT score card is maintained for each individual at the company level. The APFT consists of a 2-minute timed push-up test, a 2-minute timed sit-up test and a 2-mile timed run. In addition to these data, information regarding the individual's height and weight are also typically recorded on the APFT scorecard.

One thousand three hundred eighty three (n=1,383) APFT score cards, representing approximately 64.4% of the brigade, were initially collected in November 1994. An additional 262 APFT score cards (21.6% of the functional cohort) were discovered for individuals in the first and second battalions during one of three data collection trips between May 1995 and March 1996, yielding a total of 1645 collected APFT score cards. Of the 1214 subjects in the functional parent file, APFT score cards could be obtained on only 1019 (83.9%). The dates of these assessments, though generally available, are not included in the analyses presented here.

The 1214 subjects in the functional parent file were distributed among 10 Airborne companies, each of which apparently had a different method for recording an individual's height and weight at the time of the APFT. Some companies did record this information on the APFT score card, while others constructed separate rosters of these data. Of the 1214 subjects in the functional parent file, height and weight data could be obtained on only 799 (65.8%).

Personnel Data

The Defense Manpower Data Center (DMDC) has been building a historical archive on all active duty soldiers since 1974. This database primarily contains demographic information on each individual. Much of these data were not available at either the brigade, battalion or company level on the population being studied, but are contained in the USARIEM TAIHOD (Amoroso, 1997). Of the 1214 subjects in the functional parent file, personnel files from the TAIHOD were successfully abstracted on 1202 (99.0%).

DATA ENTRY

The medical review forms and the photocopied records from the dental questionnaires, jump logs and physical fitness score cards were entered into a pre-designed data entry system. The data from each source (i.e., dental records, medical records, etc.) were double entered into the database. After a computerized search for a subject within the parent file, a menu was utilized directing the data clerk to the appropriate data entry screen. The relevant data were entered and the data clerk returned to the parent file where a search could be conducted on the next individual. This process was continued until all data for that data source were complete. A different data entry specialist then followed the same procedure to complete the second entry of the same data. After data entry was complete twice for a specific data source, the two entries of the data were compared electronically via either EpiInfo or SAS[7]. Discrepancies between the two data entries were corrected by checking the photocopy of the original data source.

DATABASE CONSTRUCTION FOR FAILURE TIME ANALYSIS

The primary objective of this research was to investigate the risk factors for injury in the recurrent event setting. Only injuries that occurred to either the lower extremity or low back and that were musculoskeletal (not poisoning or environmental injury) were considered. In addition, we were specifically interested in investigating if an injury to a specific body part increased the risk of a subsequent injury to either the same or adjacent body part. Only the first and second injuries to the body parts of interest were included in the analysis, as there were very few subjects with three or more injuries. Two failure time models were utilized:

1) Prentice, Williams, and Peterson (PWP) Model of recurrent events in the two event setting (Prentice et al., 1981).

2) Cox Proportional Hazards Model of the time of the last injury event (Cox, 1972).

Potential explanatory variables were extracted from the dental questionnaires, physical fitness scorecards and personnel data sources. Self-reported binary data regarding cigarette and alcohol use were taken from the dental questionnaires. The continuous variables corresponding to an individual's performance in the 2-minute timed push-up test, 2-minute timed sit-up test and 2-mile timed run were extracted from the physical fitness score card. Additionally, body mass index, a measure of body density, was calculated from each individual's anthropometric data.

Using the demographic information from the TAIHOD, age at entry to the study was calculated and was used as a covariate in the Cox Model for final injury, and the first stratum (first injury event) in the PWP Model. Age at day of the first injury was also calculated when applicable and was used as a covariate in the second stratum (second injury event) in the PWP Model. A binary variable describing marital status and design variables representing ethnicity were also constructed from the TAIHOD. The referent group for ethnicity was Caucasian, and the design variables were representative of Blacks, Hispanics, and Other Ethnicity.

Selected potential explanatory variables were extracted from the medical records and were used in the second stratum (second injury event) in the PWP Model. These included type of preceding injury and highest level of medical provider seen for the preceding injury. These variables were not used in the stratum representing the first injury event because these were considered measures of the sequelae of the first injury.

A binary variable describing previous injury history during the study interval was created for the Cox Model for last injury. Thus an individual with less than or equal to one injury during the study interval was considered to have no injury history. Conversely, individuals with more than one injury were considered to have a previous injury history at the time of their second event.

ANALYSIS

Preliminary analysis included the calculation of descriptive data for both the potential explanatory variables and the outcome of interest. Means and standard deviations were calculated for all continuous variables. Frequency and relative frequency distributions were computed for all discrete variables. The number of total traumatic, overuse, and unspecified pain injuries were calculated, as well as the number of specific injury diagnoses (i.e., fracture) in each of these groups. This information was also calculated separately for the first and second injury events. Chi-square tests were performed to test the differences in the proportion of injury type and specific diagnosis between the first and second injury events. Similarly, the number of injuries to specific body parts was calculated, and chi-square tests were performed to test the differences in the proportion of affected body parts between the first and second injury events.

We hypothesized that the differing length of follow-up between the two battalions might necessitate that all regressions be stratified by battalion. Therefore, prior to model building, Kaplan-Meier estimates of the survivor function, as well as log-rank tests, were computed to determine if there were significant differences in these distributions by battalion.

A stepwise procedure was initially utilized to construct the PWP Model. This enabled the number of independent variables to be reduced to only those that may be statistically significant. The stepwise procedure implemented a p-value for entry at 0.25 and a p-value for removal at 0.80. The high p-value for removal was used so that potential confounders would not be prematurely removed from the analysis. If a design variable remained in the model after the execution of the stepwise procedure, all design variables associated with the original categorical variable were retained. Starting with the remaining independent variable with the largest Wald Chi-square p-value, variables were individually removed from the model. The log-likelihood test was implemented to determine model improvement. If the removal of a variable created a change of greater than 20% to the coefficient of another covariate, that variable was considered to be a confounder and was retained in the model. Design variables associated with a single categorical variable that were non-significant and non-confounding were removed from the model as a group. After ascertainment of the best main effects model, the scale of continuous variables was assessed using smoothed scatter plots of the Martingale residual for the model against the continuous variable of interest. Clinically plausible interactions were explored and added to the model if statistically significant.

After the best model was determined, the proportional hazards assumption was tested for each predictor in the model. The proportional hazards assumption was tested by adding a variable representing the interaction of the predictor with the logarithm of the time. Significance levels less than 0.05 tentatively suggested a violation of the proportional hazards assumption. For predictors violating the proportional hazards assumption according to this test, a log-cumulative hazard plot was conducted (Collett, 1994), a plot of the negative logarithm of the estimated survivor function against the logarithm of the survival time. In order to construct these plots for continuous variables, the variable was divided into quartiles. Near parallel curves suggested that the violation of proportionality was not severe and could be reasonably ignored.

In developing a Cox Model of the time to last injury, we sought to determine if previous injury history within the study interval was a risk factor for subsequent history. Initially, a crude hazard ratio was calculated by having only the variable representing previous injury history as a dependant variable. This hazard ratio was then adjusted with respect to explanatory variables that were significant in either strata of the PWP Model. The rationale for this alternative approach to model development was to calculate the increased risk for injury that was attributable to having a recent (within the study interval) previous injury. Additionally, the effect of previous injury history on predictors from the other models could be examined.

RESULTS

DEMOGRAPHICS

Descriptive characteristics of the study population are presented in Tables 1 and 2. This is an all-male, predominately white, physically fit, young population. The average age at entry to the study interval was approximately 24.0 (SD=5.0) years. The average performance for the 2-minute timed push-up test and the 2-minute timed sit-up test were 66.8 (SD=12.8) and 69.6 (SD=11.3) repetitions, respectively. The mean performance for the 2-mile timed run was 13.7 (SD=1.3) minutes. The average height, weight and body mass index, a measure of body density, were 1.8 (SD=0.07) meters, 76.1 (SD=9.5) kilograms and 24.5 (SD=2.6) kg/meters^2, respectively. More than half of the population reported that they were alcohol users and 30% reported that they were cigarette smokers. Approximately 38% of the population were married and 79% were Caucasian.

TABLE 1 - Descriptive Data for Age, Fitness, and Anthropometric Variables

	N	mean	SD
Age at entry to study interval (years)	1202	23.97	5.00
Push-ups (repetitions in 2 minutes)	1014	66.83	12.80
Sit-ups (repetitions in 2 minutes)	1018	69.58	11.32
Run time (minutes for 2 miles)	1011	13.69	1.32
Height (meters)	799	1.76	0.07
Weight (kilograms)	799	76.11	9.52
Body Mass Index (KG/meters^2)	799	24.48	2.57

TABLE 2 - Descriptive Data for Gender, Alcohol and Cigarette Use, Marital Status, and Ethnicity

	N	% yes
Male gender	1214	100%
Current alcohol user	1159	55.2%
Current cigarette user	1160	30.4%
Married	1201	38.0%
Ethnicity – White	946	78.7%
Black	125	10.4%
Hispanic	64	5.3%
Other	67	5.5%

INJURY

There were a total of 919 injuries during the study interval, 809 (88.0%) of which were musculoskeletal injuries; the remaining 110 (12.0%) were either environmental injuries or could not be classified in either group. Five hundred seventy five (n=575) of the 919 total injuries occurred to either the lower extremities or the low back, of which 520 (90.4%) were musculoskeletal in nature. Column I in Table 3 summarizes the number of uncensored observations available for modeling each injury event, while column II shows the distribution of the multiple events. Only 460 of the 520 (88.5%) lower extremity and low back injuries occurring as either a first or second event were of interest in this report. These injury events are denoted in bold face type in Table 3.

TABLE 3 - Injury Events Available for Analyses (I) and Distribution of Multiple Injuries (II)

Number of Injury Event	I Number of Study Subjects Experiencing this Event	II Number of Study Subjects for whom this is Event Total
0	**875**	**875**
1	**339**	**218**
2	**121**	**74**
3	47	35
4	12	11
5	1	1

Table 4 summarizes the periods of follow-up. The mean time contribution for members of first battalion was 340.8 days (SD=96.4) and ranged from 21 to 396 days. Members of second battalion contributed an average of 397.3 days (SD=165.6) and ranged from 19 to 549 days. Table 4 also summarizes the distribution of injury event numbers by battalion. For the analysis of the first injury, there were 339 events and 875 censored individuals. There were 121 events and 218 censored subjects for the analysis of the second event. The distribution of the number of injury events did not differ by battalion.

TABLE 4 - Contribution of the Analysis of Time to Event by Battalion

	1st Battalion	2nd Battalion	Total
Person Time (days)			
n	614	600	1214
mean	340.8	397.3	368.7
SD	96.4	165.6	138.0
min	21	19	19
max	396	549	549
Analysis of 1st Injury			
# of Events	170	169	339
# Censored	444	431	875
Total	614	600	1214
Analysis of 2nd Injury			
# of Events	60	61	121
# Censored	110	108	218
Total	170	169	339

Table 5 summarizes specific injury diagnosis by category (traumatic, overuse or unspecified pain) for all lower extremity and low back injuries. First and second injury events are separately displayed. The percentage of total injuries, as well as the percentage by injury category, is given for each diagnosis. Chi-square tests suggest that the proportion of first and second injury events are not statistically different in terms of either injury category or specific diagnosis.

TABLE 5 - Injury Type and Diagnosis for Total Lower Extremity/Low Back Musculoskeletal Injuries and by Event (Injury) Number

Injury Category	Diagnosis	Total n (% of total inj) (% of inj type)	1st Event n (% of total inj) (% of inj type)	2nd Event n(% of total inj) (% of inj type)
Traumatic	Sprain/Strain	203 (44.1) (70.7)	151 (44.5) (71.9)	52 (43.0) (68.4)
	Contusion	35 (7.6) (12.2)	23 (6.8) (11.0)	11 (9.1) (14.5)
	Fracture	20 (4.3) (7.0)	13 (3.8) (6.2)	7 (5.8) (9.2)
	Abrasion/Laceration	10 (2.2) (3.5)	9 (2.7) (4.3)	1 (0.8) (1.3)
	Other	19 (4.1) (6.6)	14 (4.1) (6.7)	5 (4.1) (6.6)
	TOTAL	*287 (62.4)*	*210 (61.9)*	*76 (62.8)*
Overuse	Unspecified	41 (8.9)	28 (8.3)	13 (10.7)
	Strain	(37.7)	(35.9)	(41.9)
		31 (6.7)	25 (7.4)	6 (5.0)
	Stress Fx/Rxn	(28.4)	(32.0)	(19.4)
		14 (3.0)	11 (3.2)	3 (2.5)
	Tendinitis	(12.8)	(14.1)	(9.7)
		11 (2.4)	6 (1.8)	5 (4.1)
	Other	(10.1)	(7.7)	(16.1)
		12 (2.6)	8 (2.4)	4 (3.3)
	TOTAL	(11.0)	(10.3)	(12.9)
		109 (23.0)	*78 (23.0)*	*31 (25.6)*
Pain	Unspecified	64 (13.9)	51 (15.0)	13 (10.7)
Totals		**460**	**339**	**121**

Information regarding the frequency of injuries to specific joints or muscle groups is presented in Table 6. The most common body parts injured were the ankle, knee, low back, and foot (including the toes), and together accounted for 402 (87.4%) of the 460 injuries of interest. The proportions of all injuries that occurred to the lower leg (shin and calf) were not homogeneous between the first and second injury events (Chi-square p=.026), with 24 of the 26 injuries (92.3%) occurring to this region as the first injury. There were no significant differences in the proportions of injuries between first and second injury events for all other body parts.

TABLE 6 - Body Part Affected for Total Lower Extremity/Low Back Musculoskeletal Injuries and by Event (Injury) Number

Body Part Affected	Total n (% of total inj)	1st Event n (% of total inj)	2nd Event n (% of total inj)
Ankle	112 (24.3)	88 (26.0)	24 (19.8)
Knee	105 (22.8)	72 (21.2)	33 (27.3)
Low Back	95 (20.7)	71 (20.9)	24 (19.8)
Foot/Heel/Toe	90 (19.6)	61 (18.0)	29 (24.0)
* Shin/Calf	26 (5.7)	24 (7.0)	2 (1.7)
Hip	17 (3.7)	10 (2.9)	7 (5.8)
Thigh	11 (2.4)	9 (2.7)	2 (1.7)
Multiple Low	4 (0.9)	4 (1.2)	0 (0.0)
Totals	**460**	**339**	**121**

* denotes a significant difference ($p<.05$) of the proportion of affected body parts by event

FAILURE TIME REGRESSION ANALYSES

Prior to model building, Kaplan-Meier estimates of the survivor function, as well as log rank tests, were conducted to ascertain that there was no difference between the two battalions. Both the Kaplan-Meier estimates, as well as the log rank tests, revealed no significant differences between the two battalions (data not shown). This verified that regression models could be constructed by stratifying by battalion, therefore accommodating for the different length of the retrospective medical record reviews.

The Prentice Williams and Peterson (PWP) Model

The first strata (first injury event) of the PWP Model resulted in parameter estimates that suggest that increased hazard of injury is associated with lower push-up performance, lower sit-up performance, and younger age at entry to study. Also predictive of increased hazard were use of alcohol and being married (Table 7). Specifically, a 10-unit decrease in upper body strength and endurance as measured by the 2-minute timed push-up test was associated with a 16.2% increased risk of lower extremity musculoskeletal injury ($P < .01$). Similarly, a 10-unit decrease in abdominal and hip flexor strength and endurance as measured by the 2-minute timed sit-up test resulted in a 15.2% increased risk of injury ($P < .05$). A 1-year increase in age resulted in a 4.1% decrease in risk of lower extremity injury ($P < .01$). Alcohol users were at 30.4% greater risk of injury ($P < .05$). There is evidence that married subjects had a relative odds of injury 27.0% greater than that of their unmarried counterparts ($P < .10$). This variable was retained in the final model primarily because of its marginal significance and that it was a confounder on the variable, age, at entry to study.

Removal of the marital status variable from the model resulted in a 32.0% change in the parameter estimate for age (data not shown). Previous research links tobacco use and injury risk (Amoroso, 1996). In this analysis, tobacco use fell out of the model, even if alcohol use was forced out first.

The second strata (second injury event) of the PWP Model resulted in parameter estimates that suggest that the following were associated with a repeat injury: decreased push-up performance, traumatic first injury, and seeing only a medic as the highest level of provider for the first injury. Additionally, Hispanics were at increased risk and individuals in the "Other Ethnicity" category were at lower risk (Table 7). A 10-unit decrease in upper body strength and endurance as measured by the 2-minute timed push-up test resulted in a 24.9% increased risk of subsequent injury ($P < .01$). If the subject's first injury was categorized as a traumatic injury, there was an 83.4% increased risk of subsequent injury than if the first injury was categorized as overuse or unspecified pain ($P < .01$). Subjects who saw only a Medic, the lowest level of medical provider, for the preceding injury were 71.6% more likely to undergo a subsequent injury ($P < .05$). Hispanic individuals had greater than four times the risk of experiencing a second lower extremity injury than did Caucasian individuals ($P < .001$). Two thirds of the Hispanic subjects who were at risk to experience a second lower extremity injury did so, compared to 34.3% of the remainder of the subjects (data not shown).

TABLE 7 - Parameter Estimates for Final PWP Model

Stratum=1 (1st Injury)	Parameter Estimate	SE	Hazard Ratio (95% CI)
Push-ups (10 repetition decrease)	0.150 ***	0.056	1.162 (1.042, 1.296)
Sit-ups (10 repetition decrease)	0.142 **	0.063	1.152 (1.018, 1.305)
Age at entry to study (1 year increase)	-0.040 ***	0.015	0.961 (0.933, 0.989)
Alcohol user (vs. abstainer)	0.265 **	0.121	1.304 (1.028, 1.653)
Married (vs. non-married)	0.239 $*	0.141	1.270 (0.963, 1.675)
Stratum=2 (2nd Injury)			
Push-ups (10 repetition decrease)	0.223 ***	0.083	1.249 (1.062, 1.470)
Previous Traumatic Injury	0.607 ***	0.217	1.834 (1.200, 2.804)
Highest level of Medical Provider from Previous Injury: Medic (vs. all others)	0.540 **	0.251	1.716 (1.049, 2.808)
Ethnicity (referent = White): Black Hispanic Other	0.037 1.446 **** -1.243 *	0.374 0.368 0.726	 4.246 (2.023, 8.738) 0.289 (0.070, 1.120)

Model Chi-Square = 69.289, 11df (p=0.0001)
*P< .1 **P< .05 ***P< .01 ****P< .001 $confounded with age (1st stratum)

The Cox Model of Time to Last Injury

The purpose of the Cox Model of time to last injury was to determine the magnitude of the increased hazard associated with a previous injury event. Table 8 shows both crude and adjusted values of parameter estimates, standard errors and hazard ratios for the history of previous injury, as well as for significant variables in the PWP Model. The adjusted value is adjusted for variables that were statistically significant in the PWP Model, which included push-up performance, sit-up performance, age at entry to study, alcohol user (vs. abstainer), marital status and ethnicity. The variables describing previous traumatic injury and highest level of provider for the previous injury were not included in this model. This is because the last injury event is not the second injury event for all individuals. The crude and adjusted parameter estimate values corresponding to previous injury history did not differ considerably, suggesting that its effect is independent of those of the other predictors of recurrent injury and that individuals with a history of one injury are at approximately seven times greater risk of a second injury.

The effect of previous injury history is perhaps easier to interpret by examination of a log-rank test for equality of the survivor functions between those with a prior injury history versus those without a prior injury history. Table 9 illustrates that if previous injury history was not a risk factor for subsequent injury, only 25 of the 339 injury events would have been a second injury event. The actual number of second injury events (individuals with a previous injury history) was 121, a value that is almost five times greater than the expected number of subjects with previous injury history.

TABLE 8 - Crude and Adjusted Parameter Estimates for History of Previous Injury and Other Covariates from Cox Regression Model for Last Injury

	Parameter Estimate	SE	Hazard Ratio (95% CI)
History of Previous Injury (crude)	**2.005** *****	**0.117**	**7.426 (5.905, 9.338)**
(adjusted)	**1.941** *****	**0.130**	**6.965 (5.394, 8.992)**
Push-ups (10 rep decrease) (crude)	0.204 ****	0.046	1.227 (1.121, 1.342)
(adjusted)	0.058	0.055	1.060 (0.951, 1.181)
Sit-ups (10 rep decrease) (crude)	0.207 ****	0.052	1.230 (1.110, 1.363)
(adjusted)	0.156 **	0.065	1.169 (1.030, 1.327)
Age (1 yr. increase) (crude)	-0.033 ***	0.012	0.968 (0.945, 0.990)
(adjusted)	-0.038 **	0.015	0.963 (0.935, 0.991)
Alcohol user (vs. abstainer) (crude)	0.107	0.111	1.113 (0.895, 1.385)
(adjusted)	0.290 **	0.123	1.337 (1.051, 1.701)
Married (vs. non-married) (crude)	-0.022	0.112	0.978 (0.785, 1.218)
(adjusted)	0.080	0.139	1.083 (0.824, 1.423)
Ethnicity : Black: (crude)	-0.093	0.185	0.911 (0.635, 1.308)
(adjusted)	0.244	0.221	1.277 (0.828, 1.968)
Hispanic: (crude)	-0.164	0.265	0.849 (0.505, 1.428)
(adjusted)	-0.125	0.298	0.883 (0.492, 1.582)
Other: (crude)	-0.122	0.258	0.885 (0.534, 1.466)
(adjusted)	0.209	0.289	1.232 (0.700, 2.170)

Model Chi-Square for history of previous injury (crude) = 228.15, 1df (p=0.0001)
Model Chi-Square for history of previous injury (adjusted) = 220.44, 9df (p=0.0001)
*P< .1 **P< .05 ***P< .01 ****P< .001 ***** P< .0001

TABLE 9 - Log-Rank Test for Equality of Survivor Functions for Previous Injury History

	Observed	Expected
No previous injury history	218	314
Previous injury history	121	25
Totals	339	339

Chi-square = 400.5, 1df ($p < 0.0001$)

Proportional Hazards Assumption

An underlying assumption of the Cox Model is that the survival time among individuals in two or more different groups of a significant variable are proportional to one another. Since the PWP Model reduces to the Cox Model in the absence of more than one event, it is reasonable to examine this assumption for this model as well. One method of testing this assumption is by adding a covariate to the final model that is representative of the interaction between the covariate of interest and the logarithm of the time variable. If this interaction term has a corresponding small P-value (< .05), one would conclude that the survival time between individuals with different values of this covariate are not proportional. In other words, the effect of this covariate is not the same at all points in time.

Variations of the effect of a covariate with time may not, however, be of concern. The parameter estimates for a significant covariate represents the average effect of that covariate over the range of time observed in the data (Allison et al., 1995). Apparent violations of the proportional hazards assumption were therefore checked via log-cumulative hazard plots for each of the above models. This is a plot of the negative logarithm of the estimated survivor function on the vertical axis against the logarithm of the time variable on the horizontal axis (Collett, 1994). Figures 1, 2, and 3 are the log-cumulative hazard plots for all significant variables in the first stratum of the PWP Model, the second stratum of the PWP Model, and the Cox Model to each individual's last injury, respectively. Variables representing quartiles were used to construct plots for continuous variables. Only plots representing the significant covariates in the adjusted model are shown for the Cox Model to an individual's last injury (Figure 3).

Notice that the different strata in each log-cumulative hazard plot do not vary greatly from one another. The small divergences between the different curves on each plot represent that the survival times between individuals in these different groups may not be perfectly proportional; however, this also exhibits that this non-proportionality is not of concern in this setting.

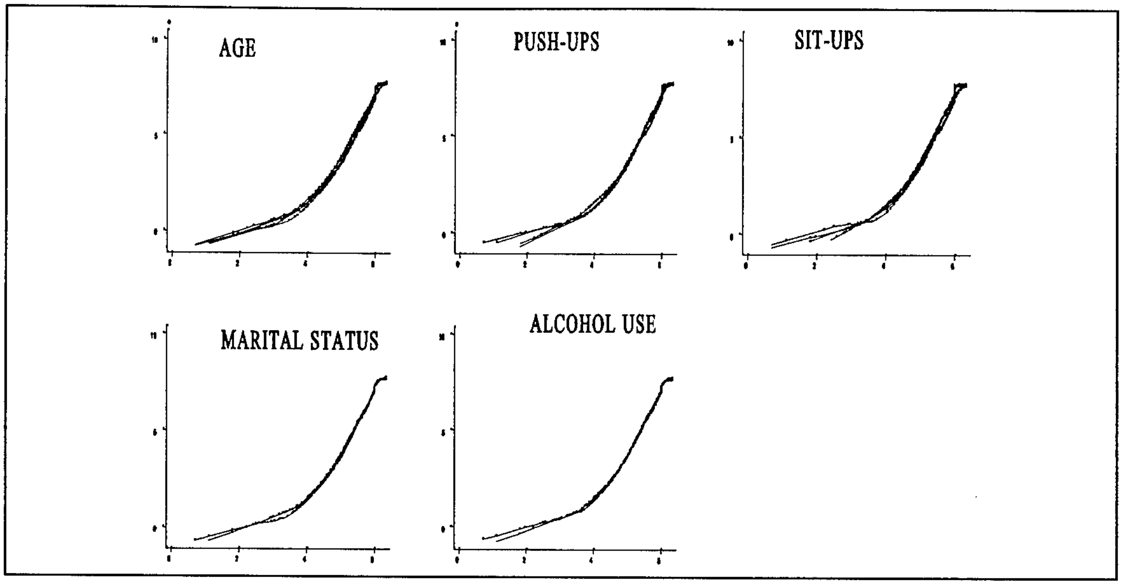

Figure 1 - Log-Cumulative Hazard Plots for First Stratum of PWP Model

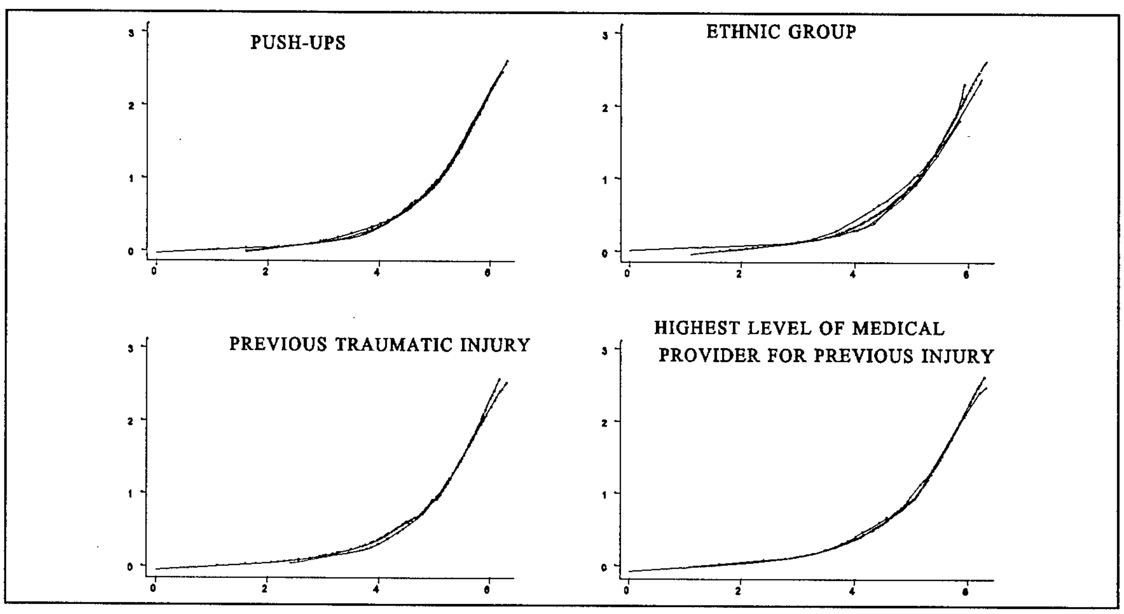

Figure 2 - Log-Cumulative Hazard Plots for Second Stratum of PWP Model

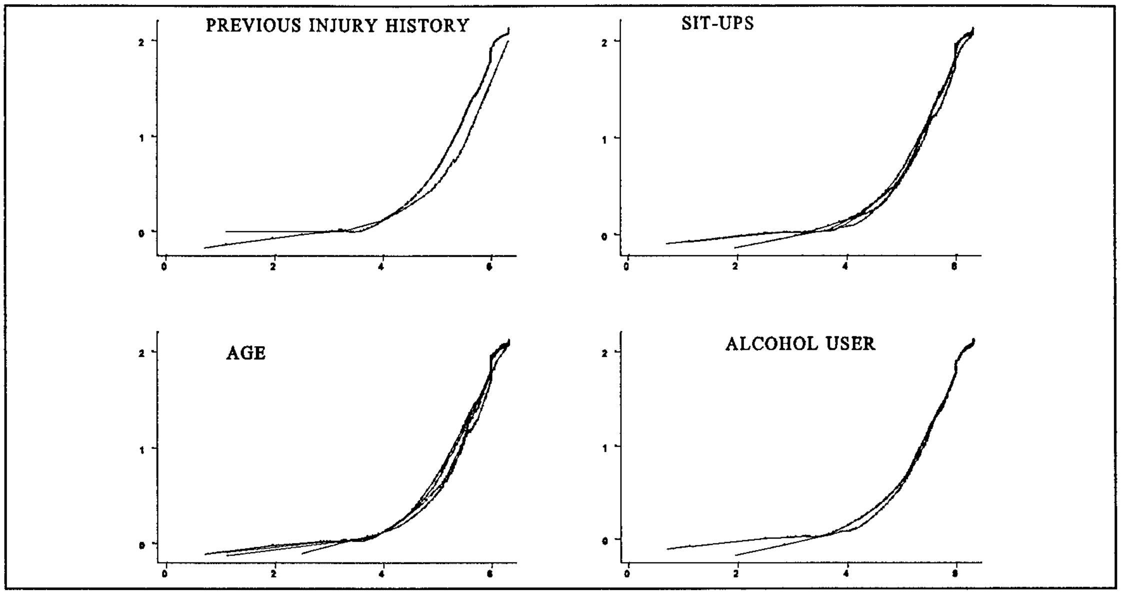

Figure 3 - Log-Cumulative Hazard Plots for Cox Model to Last Injury

DISCUSSION

The first strata of the PWP Model yielded risk factors that were similar to those customarily seen in previous injury epidemiology studies on military populations (Jones and Hanson, 1996). Marital status is a risk factor that has not previously been investigated. The fact that this analysis suggests that married soldiers are 27% more likely to sustain an injury than their unmarried counterparts ($p < .10$) warrants further examination. The risk factors associated with the second strata of the PWP Model, with the exception of upper body strength and endurance as measured by the 2-minute timed push-up test, are very different from the commonly agreed upon risk factors for injury in military populations. There is a highly significant difference in the risk of a subsequent injury among different ethnic groups which should be further examined. However, the risk factors associated with the sequela of the preceding injury are perhaps of greater interest. Specifically, the fact that this analysis suggests that individuals whose immediately preceding injury was a traumatic event (as opposed to an overuse or an unspecified pain injury) have an 83% increased likelihood to sustain a subsequent injury suggests that there should perhaps be a change in the medical management of these individuals. For example, one could hypothesize that this increased risk is a result of inadequate recovery time after a traumatic injury. If future research proves this to be true, it would therefore be reasonably simple to minimize this excess risk.

Another aspect of the sequela of the previous injury that was a significant risk factor for subsequent injury was the highest level of medical provider seen for the immediately preceding injury. Specifically, those individuals who saw only a medic, the lowest level of medical provider, after their initial injury were at a 72% increased risk of enduring a subsequent injury. This is likely because medics are the only medical providers that do not have the authority to order a restriction on an individual's activity. This should be further examined, and if proven to be true, simple changes in the medical management of injured individuals may provide a simple injury control measure.

While it is intuitive that injured individuals are at increased risk for a subsequent injury, especially in the time immediately following the event, previous injury epidemiology has predominately ignored this phenomena. The Cox Model of time to last injury supplies information regarding the magnitude of the increased risk associated with having a recent previous injury history. These data show that having a recent injury increases an individual's risk of having a subsequent injury approximately seven-fold. Additionally, this increased risk is independent of the other significant risk factors. It is also possible that this seven-fold increase in risk associated with previous injury history is an underestimation of the truth. This analysis examined only the first and second injury events for an interval that did not exceed 18 months. Therefore, if the increased risk associated with recent previous injury is cumulative, examination of this occurrence over a longer time duration and for a larger number of injuries might be expected to show a still greater risk associated with previous injury history. A limitation of this analysis is that it assumes individuals had no prior injury history prior to the beginning of this study. This is not likely to be true, therefore suggesting that the increased risk associated with previous injury history may subside with time. This possibility should also be examined in greater detail.

BENEFITS OF THE PWP MODEL

The PWP Model has been determined to be the most appropriate failure time model to analyze these data and perhaps, more generally, data in the multiple injury event setting (Schneider, In Press). This model has the benefit of permitting a different baseline hazard function for each event, thus allowing for the associated risk of a particular covariate to be compared across strata (injury events). Table 10 displays the hazard ratios for the final PWP Model, as well as for a "comparison" model that includes, in both strata, all covariates that were significant in either stratum. In this manner, changes in the parameter estimates from one injury to the next can be easily compared. Table 10 also shows the percentage of change in the parameter estimates (the natural logarithm of the hazard ratio) between the final and comparison model where applicable. There was no significant confounding of parameter estimates between the two models.

A practical advantage of the PWP Model is inherent within the structure of the database. Specifically, Kaplan-Meier survivor estimates to the time of injury can easily be plotted on the same axis. Figure 4 shows that the estimated survivor function for

TABLE 10 - Parameter Estimates for Final and "Comparison" PWP Models

	Hazard Ratio for Final Model	Hazard Ratio for "Comparison Model"	% Change in Parameter Estimate
Push-up stratum=1 Push-up stratum=2	1.162*** 1.250***	1.160*** 1.310***	-1.35 % 17.41 %
Sit-up stratum=1 Sit-up stratum=2	1.153** NA	1.154** 1.070^{ns}	0.70 % NA
Age stratum=1 Age stratum=2	0.961*** NA	0.962*** 0.977^{ns}	-2.56 % NA
Alcohol stratum=1 Alcohol stratum=2	1.303** NA	1.297** 0.816^{ns}	-1.92 % NA
Marital Status stratum=1 Marital Status stratum=2	1.270$^{\$*}$ NA	1.273* 1.189^{ns}	0.83 % NA
Ethnicity (referent=white) Black stratum=1 Black stratum=2 Hispanic stratum=1 Hispanic stratum=2 Other stratum=1 Other stratum=2	 NA 1.038^{ns} NA 4.246**** NA 0.289*	 0.982^{ns} 1.035^{ns} 0.782^{ns} 4.618**** 0.876^{ns} 0.289*	 NA -8.82 % NA 5.49 % NA -0.24 %
Prev traum injury stratum=1 Prev traum injury stratum=2	NA 1.835***	NA 1.793***	NA -3.94 %
Prev Provdr- medic stratum=1 Prev Provdr- medic stratum=1	NA 1.716**	NA 1.844**	NA 11.76 %

*P< .1 **P< 05 ***P< .01 ****P< .001 ^{ns}P > .1
NA = Not Applicable $^{\$}$confounded with age

the first injury is consistently less than that of the second injury. This suggests that once an individual experiences a traumatic injury to the lower extremity or low back, he is at greater risk to undergo a similar subsequent injury. The results of the log-rank for the equality of the survivor function between the two strata in the PWP Model are shown in Table 11. This specifies that of the 460 injuries of interest, 55 of them are expected to be an individual's second injury; however, the actual number of second injuries is 121, more than twice the expected number. This yields information that is similar to that from log-rank test conducted on the binary variable previous injury in conjunction with the Cox Model to last injury (refer to Table 9).

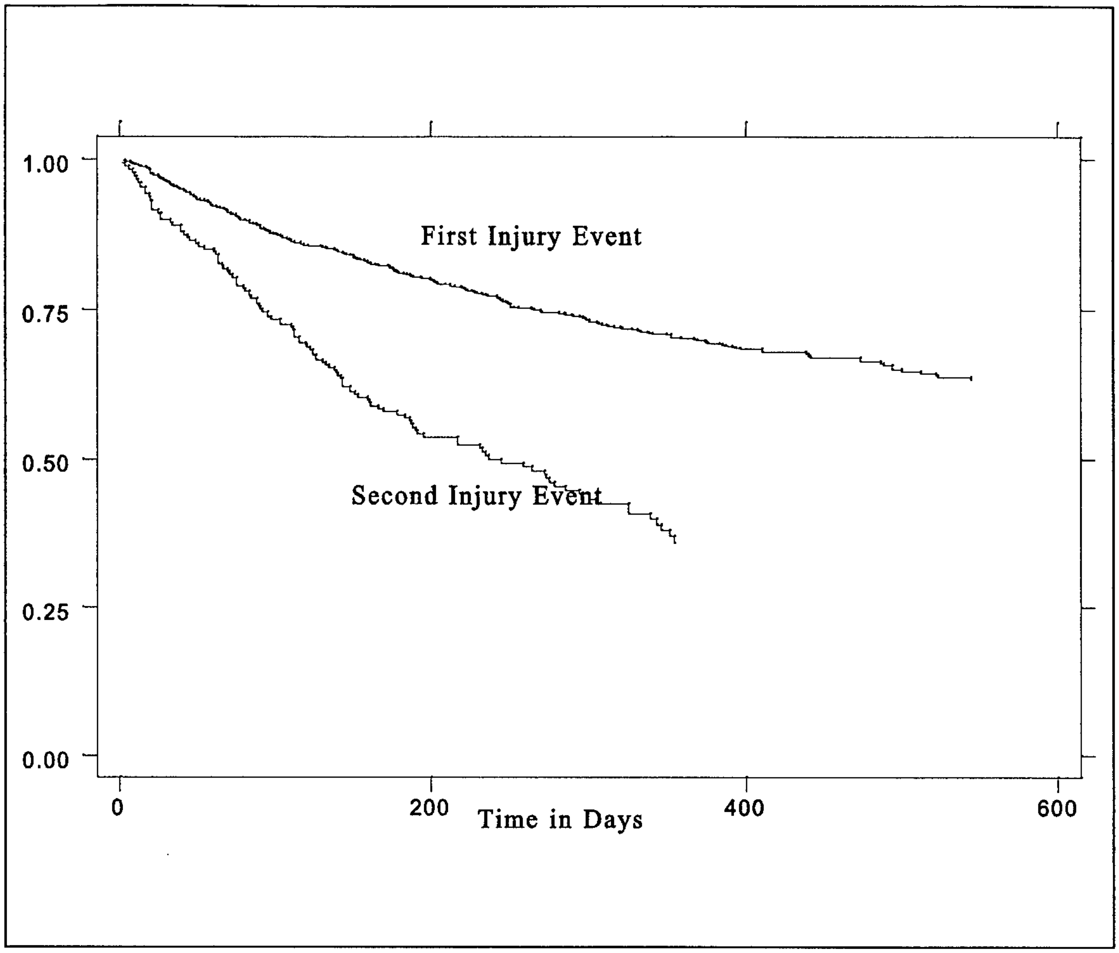

Figure 4 - Kaplan-Meier Survivor Estimates for the Two Strata (Injury Events) in the PWP Model

TABLE 11 - Log-Rank Test for Equality of Survivor Functions between the Two Strata (Injury Events) in the PWP Model

	Observed	Expected
Stratum 1 (first injury)	339	405
Stratum 2 (second injury)	121	55
Totals	460	460

Chi-square = 89.49, 1df ($p < 0.0001$)

BENEFITS OF THE COX MODEL OF THE TIME TO LAST INJURY

While the PWP Model is the best for analyzing recurrent injuries, it does have one limitation. Specifically, the increase in hazard attributable to having a previous injury is not directly estimable. The PWP Model does allow for verification that there is an increased risk for injury having had a previous injury by way of Kaplan-Meier estimates of the survivor function and log-rank tests of the different strata (injury events). However, the Cox Model of the time to the last injury event is needed to fully understand the effect of previous injury history on subsequent injury, as this model estimates the increased hazard associated with having had a previous injury.

REFERENCES

Allison, P.D. Survival Analysis Using the SAS[7] System: A Practical Guide. SAS Institute Incorporated, Cary, NC, 1995.

Amoroso, P.J., K.L. Reynolds, J.A. Barnes, and D.J. White. Tobacco and injury: An annotated bibliography. USARIEM Technical Note TN96-1, 1996.

Amoroso, P.J., W.G. Swartz, F.A. Hoin, and M.M.Yore. Total Army Injury and Health Outcomes Database: Description and capabilities. USARIEM Technical Report TN97-2, 1997.

Andersen, P.K., O. Borgan, R.D. Gill, and N. Keilding. Statistical Models Based on Counting Processes. Springer-Verlag, New York, 1993.

Andersen, P.K. and R.D. Gill. Cox's Regression Model for counting processes: a large sample study The Annals of Statistics, 10:1100-1120, 1982.

Brodine, S.K. and R.A. Shaffer. Unpublished data, 1995.

Collett, D. Modeling Survival Data in Medical Research. Chapman and Hall, New York, 1994.

Cox, D.R. Regression models and life tables (with discussion). Journal of the Royal Statistical Society, Series B, 34:187-220, 1972.

Hosmer, D. Survival Analysis in Epidemiology. Lectures given at New England Epidemiology Institute, Tufts University, Massachusetts, 24-28 June.

Jones, B.H., D.N. Cowan, J.P. Tomilson, D.W. Polly, and P.N. Frykman. Epidemiology of injuries associated with physical training among young men in the Army. Med Sci Sports and Exerc, 25:197-203, 1993.

Jones, B.H. and B.C. Hanson. Injuries in the Military - A Hidden Epidemic, A report by the Armed Forces Epidemiology Board Injury Work Group. United States Army Center for Health Promotion and Preventive Medicine, Aberdeen Proving Ground, MD, 1996.

Jones, B.H., J.H. Manikowski, J.H. Harris, J. Dziados, S. Norton, T. Ewart, et al. Incidence of and risk factors for injury and illness among male and female Army basic trainees. USARIEM Technical Report T19-88, 1988.

Prentice, R.L., B.J. Williams, and A.V. Peterson. On the regression analysis of multivariate failure time data. Biometrica, 68:373-379, 1981.

Province, M.A. The effects of exercise on falls in elderly patients. JAMA, 273:1341-1347, 1995.

Schneider, G.A. Multiple event analysis of injuries using adaptations to the Cox Proportional Hazards Model. USARIEM Technical Note, TN98-1, Sep 97 (ADA332905).

APPENDIX 1 - TIME DEPENDANT SINGLE EVENT ANALYSIS USING THE COX PROPORTIONAL HAZARDS MODEL

DERIVATION OF THE LIKELIHOOD

The unconditional full likelihood for a time dependent single event analysis represents each outcome as a triplet of the form

$$[t_i, c_i, \underline{x}_i(t_i)] \tag{1}$$

where

t_i = time to last contact
c_i = event indicator (0=censored, 1=event)
$\underline{x}_i(t_i)$ = vector of explanatory variables, possibly a function of t_i

with i=1 to N indexing the subjects, assumed independent.

Consider a censored outcome at time t. All that is known about this individual is that their survival time is greater than t. Thus, his/her contribution to the unconditional full likelihood is the probability that an individual with associated covariate pattern $\underline{x}$ survives until at least time t. This is synonymous with the survivorship function, S(∃). Since c=0 for a censored outcome, the contribution to the likelihood may be expressed as follows:

$$[S(t, \underline{\beta}, \underline{x}(t))]^{1-c}, \tag{2}$$

where S(∃) is the survivorship function and is assumed to be related to the vector of covariates, $\underline{x}(t)$, through an associated vector of regression coefficient $\underline{\beta}$.

Consider next an actual event occurring at time t. Here, the contribution to the likelihood is identically the density function, f(∃). Thus with c=1 for an actual event, the contribution to the likelihood function becomes the following:

$$[f(t, \underline{\beta}, \underline{x}(t))]^c \qquad (3)$$

Again, f(∃) is assumed to be related to the vector of covariates, $\underline{x}(t)$, through an associated vector of regression coefficients.

For a sample of N individuals, the unconditional full likelihood is the product of the N independent contributions. Thus,

$$L(\underline{\beta}) = \vartheta_{i=1 \text{ to } N} [S(t, \underline{\beta}, \underline{x}(t))]^{1-c} [f(t, \underline{\beta}, \underline{x}(t))]^c \qquad (4)$$

This unconditional full likelihood is completely general. No assumptions have been made about the form of S(∃) and f(∃), the link between $\underline{x}(t)$ and S(∃) and f(∃), nor the relationship between the event and censoring mechanisms. Thus, further assumptions and a model are needed in order to make inferences about β (Hosmer, 1996).

HEURISTIC OF THE COX PROPORTIONAL HAZARDS MODEL

Cox's (1972) formulation of the Proportional Hazards Model derives from a partial likelihood function that conditions on the set of actual event times and exploits two assumptions:

(1) The censoring mechanism is independent of the event mechanism.
(2) The hazard function $h(t, \underline{\beta}, \underline{x})$ is linked to the explanatory variables via the model

$$h(t, \underline{\beta}, \underline{x}) = h_o(t) \exp[\underline{x}(t)'\underline{\beta}] \qquad (5)$$

where $h_o(t)$ is an arbitrary baseline hazard function that is independent of $\underline{x}(t)$ for all t.

The advantage to conditioning on the set of actual event times is that it avoids having to make an assumption about the form of S(∃) and f(∃); e.g., exponential, Weibull, etc. The advantage to assuming that the censoring mechanism is independent of the event mechanism is that it permits analysis of a partial likelihood, which considers only observed actual events and the associated risk sets.

THE COX PROPORTIONAL HAZARDS MODEL AND THE UNDERLYING PARTIAL LIKELIHOOD

As indicated above, the conditional likelihood used to derive the partial likelihood that underlies the Cox Model conditions on the set of ordered occasions on which actual events occurred. Suppose there are n actual events, n#N. If these are denoted using the usual order statistic notation, $\{t_{(i)}\}$, the conditional likelihood of interest is that of

$$[t_1, c_1, \underline{x}_{(1)}(t_1)], \ldots\ldots, [t_n, c_n, \underline{x}_{(n)}(t_n)] * \{t_{(i)}\} \tag{6}$$

Without loss of generality, suppose C_1 individuals are censored prior to the first event, an additional C_2 individuals are censored prior to the second event, and so forth. Arguing conditionally on the set of event times $\{t_{(i)}\}$ allows us to write the conditional likelihood as the following product:

$$L_{cond}([t_1, c_1, \underline{x}_1(t_1)], \ldots\ldots, [t_n, c_n, \underline{x}_{(n)}(t_n)] * \{t_{(i)}\}) =$$

$$\begin{aligned}
&L[C_1 \text{ censored in } (0,t_{(1)}^-), \text{ No events in } (0,t_{(1)}^-)] \; \exists \\
&L[\text{1st event at } t_{(1)} * \text{ history to } t_{(1)}^-] \; \exists \ldots\ldots \exists \\
&\qquad L[n^{th} \text{ event at } t_{(n)} * \text{ history to } t_{(n)}^-] \; \exists \; L[C_{n+1} \text{ censored after } t_{(n)}]
\end{aligned} \tag{7}$$

where for ease of notation, "history to $t_{(1)}^-$," is shorthand for "no events in $(0, t_{(1)}^-)$," "history to $t_{(2)}^-$," is shorthand for "no events in $(0, t_{(1)}^-)$, one event at $t_{(1)}$, no events in $t_{(1)}^+, t_{(2)}^-$)," etc. Following Andersen et al. (1993, page 49), these "histories" are denoted $\setminus_t$. Thus $\setminus_t$ represents the available data at time t, and $\setminus_{t^-}$ represents the available data at time t^-. Thus, the conditional likelihood can be constructed incrementally over the occasions of censoring and event times by exploiting the theorem of total probabilities. The first term represents the likelihood of the first C_1 censoring, the second term represents the likelihood of the first event at time $t_{(1)}$, conditional on the history to time $t_{(1)}^-$, and so on. This is analogous to the lifetable approach to estimating survival probabilities. Further inspection reveals that this conditional likelihood contains two types of terms: one corresponding to the occasions of censoring and the other corresponding to occasions of the actual events. When these are regrouped, the full conditional likelihood is seen to be of the following form:

$$L_{cond}\{[t_1, c_1, \underline{x}_1(t_1)], \ldots\ldots, [t_n, c_n, \underline{x}_n(t_n)] * \{t_{(i)}\}\} =$$

$$\begin{aligned} &\vartheta_{j=1 \text{ to } n} L[j^{th} \text{ event at time } t_{(j)} \mid \setminus t_{(j)}^-] \\ \exists\ &\vartheta_{j=1 \text{ to } n} L[C_j \text{ censored in } (t_{(j-1)}^+,\ t_{(j)}^- * \setminus t_{(j-1)}] \\ \exists\ &L[C_{n+1} \text{ censored at } t_{(n)} * \setminus t_{(n)}] \end{aligned} \tag{8}$$

Here the subscript j indexes the n actual event times and by definition, $t_{(0)}^+ = 0$.

The Cox Model partial likelihood is extracted from the conditional likelihood by dropping the censoring likelihood terms. Justification is the assumption that the censoring mechanism is independent of the event mechanism and that the censoring likelihood terms contain no information about $\underline{\beta}$. We then obtain:

$$L\{[t_1, c_1, \underline{x}_1(t_1)], \ldots\ldots, [t_N, c_N, \underline{x}_N(t_N)] * \{t_{(i)}\}\}_{partial}$$
$$= \vartheta_{j=1 \text{ to } n}\, L[j^{th} \text{ event at time } t_{(j)} * \backslash t_{(j)}^{-}] \qquad (9)$$

The conditional likelihood, $L[j^{th}$ event at time $t_{(j)} * \backslash t_{(j)}^{-}]$, can then be seen to be equal to the following:

$$L[j^{th} \text{ event at time } t_{(j)} * \backslash t_{(j)}^{-}] =$$
$$L[j^{th} \text{ event at time } t_{(j)} * j^{th} \text{ survives to time } t_{(j)}^{-}] \;)$$
$$3_{u:\,Rj}\, L[u^{th} \text{ event at time } t_{(j)} * u^{th} \text{ survives to time } t_{(j)}^{-}] \qquad (10)$$

where R_j is the subset remaining at risk at time $t_{(j)}^{-}$.

Notice that the assumption of independence of the event and censoring mechanisms permits replacing the condition " history to time $t_{(j)}^{-}$ " with the risk set at time $t_{(j)}^{-}$, which is denoted R_j. Finally, noting that $L[u^{th}$ event at time $t_{(j)} * u^{th}$ survives to time $t_{(j)}^{-}]$ is by definition the hazard function, yields the following partial conditional likelihood:

$$L\{[t_1, c_1, \underline{x}_1(t_1)], \ldots\ldots, [t_N, c_N, \underline{x}_N(t_N)] * \{t_{(i)}\}\}_{partial} =$$
$$\vartheta_{j=1 \text{ to } n}\, [h(t_{(j)}, \beta, \underline{x}_j) \;)\; 3_{u:\,Rj}\, h(t_{(j)}, \beta, \underline{x}_u)] \qquad (11)$$

As in lifetable methods of estimation, censored observations are retained in the likelihood as long as they are at risk.

Finally, substitution of the Cox Model assumption yields:

$$L\{[t_1, c_1, \underline{x}_1(t_1)], \ldots\ldots, [t_n, c_n, \underline{x}_n(t_n)] * \{t_{(i)}\}\}_{partial} =$$
$$\vartheta_{j=1 \text{ to } n}\, \{\, h_o(t_{(j)}) \exp[\underline{x}(t_{(j)})'\underline{\beta}] \;)\; 3_{u:\,Rj}\, h_o(t_{(j)}) \exp[\underline{u}(t_{(j)})'\underline{\beta}] \,\} =$$
$$\vartheta_{j=1 \text{ to } n}\, \{\exp[\underline{x}(t_{(j)})'\underline{\beta}] \;)\; 3_{u:\,Rj}\, \exp[\underline{u}(t_{(j)})'\underline{\beta}] \,\} \qquad (12)$$

Maximum likelihood based inference for $\underline{\beta}$ is based on this function.

APPENDIX 2 - TIME DEPENDANT MULTIPLE EVENT ANALYSIS USING THE PRENTICE, WILLIAMS AND PETERSON (PWP) MODEL

The PWP Model is an alternative to the Cox Model that has less stringent assumptions than other multiple event regression techniques. Specifically, it allows both the baseline hazard and the values of the regression parameters to vary by event. This is accomplished through the use of stratification. Strata are defined according to the number of previous events (Prentice, Williams, and Peterson, 1981).

As this report concerns the modeling of one recurrence of event, the number of previous events can be only 0 or 1. Let s = 0,1 index the number of preceding events, thus indicating the strata. The PWP Model formulates the intensity process, hence the change in the counting process over a small increment of time, separately for each stratum s:

$$\lambda_{sj}(t) = Y_{sj}(t)\, h_{s0}(t)\, \exp[\underline{x}_j(t)'\underline{\beta}_s] \tag{13}$$

Formulation of the PWP Model permits a separate such linking for each number of preceding events. Thus, the baseline hazard of an event varies depending on the number of preceding events. As well, the effect of the covariate pattern history can also vary with respect to the number of preceding events. Let s=0,1,......S index the number of preceding events. (Note: In this report, where interest is in the analysis of one recurrence of injury, s is either 0 or 1.)

The PWP Model further allows for time zero to be defined in various ways depending on the interest of the investigator. Prentice, Williams and Peterson suggest two definitions for time zero: the time t since the beginning of the study, and the time t-$t_{n(t)}$, the time since the immediately preceding event, which is often referred to as the gap-time model. Thus, the PWP Model formulates the instantaneous risk of an event at time t as a function of the number of events history and the covariate pattern history as follows:

$$h(t\,,\,\beta,\,x(t) * n(t)=s) = h_{0s}(t)\exp(\underline{x}(t)'\underline{\beta}_s) \quad \text{and} \tag{14}$$
$$h(t\,,\,\beta,\,x(t) * n(t)=s) = h_{0s}(t-t_{n(t)-1})\exp(\underline{x}(t)'\underline{\beta}_s) \tag{15}$$

for the time since the beginning of the study and the time since the immediately preceding event, respectively, where,

$s=0,1\ldots S$ = the number of preceding events
$h_{0s}(t)$ and $h_{0s}(t-t_{n(t)-1})$ = the corresponding baseline hazard functions for the two possible time scales
$\underline{\beta}_s$ = the vector of stratum specific regression coefficients.

The formulation of the PWP Model is most easily understood in the context of the derivation of the Cox Model. Recall first the Cox Model's link of the hazard function, h(∃), to the explanatory variables $\underline{x}$ and the associated regression parameters $\underline{\beta}$:

$$h(t,\,\underline{\beta},\,\underline{x}) = h_o(t)\ \exp[\underline{x}(t)'\underline{\beta}] \tag{16}$$

Recall next that the partial likelihood for the Cox Model is defined:

$$L_{partial} = \vartheta_{j=1\text{ to }n}\{\exp[\underline{x}(t_{(j)})'\underline{\beta}]\ )\ 3_{u:\,Rj}\ \exp[\underline{u}(t_{(j)})'\underline{\beta}]\ \} \tag{17}$$

The PWP Model adapts this partial likelihood, employing the stratification of number of previous events, where all individuals in a given strata are homogeneous with respect to the number of preceding events. The corresponding partial likelihood equation for the time since the beginning of the study is therefore:

$$L_{PWP1} = \prod_{s=0 \text{ to } S} \prod_{k=1 \text{ to } ds} \{\exp[\underline{x}_{\underline{sk}}(t_{sk})'\underline{\beta}_{\underline{s}}] \,/\, \sum_{u \in R(tsk,\, s)} \exp[\underline{x}_{\underline{u}}(t_{sk})'\underline{\beta}_{\underline{s}}]\} \qquad (18)$$

where,

d_s = number of actual events occurring in the s^{th} stratum defined by the number of preceding events.
$R(t_{sk})$ = the subset at risk in the s^{th} stratum just prior to time t_{sk}.

For the gap-time choice of time scale, the PWP partial likelihood is expressed:

$$L_{PWP2} = \prod_{s=0 \text{ to } S} \prod_{k=1 \text{ to } ds} \{\exp[\underline{x}_{\underline{sk}}(t_{sk})'\underline{\beta}_{\underline{s}}] \,/\, \sum_{u \in R(vsk,\, s)} \exp[\underline{x}_{\underline{u}}(\tau_u + v_{sk})'\underline{\beta}_{\underline{s}}]\} \qquad (19)$$

where,

τ_u = the last failure time on subject u prior to entry into stratum s.
v_{sk} = the gap time from the immediately preceding event.

Maximum likelihood based inference for $\underline{\beta}_{\underline{s}}$ are based on these partial likelihood equations.

DISTRIBUTION LIST

1 Copy:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Fort Belvoir VA 22060-6218

Office of the Assistant Secretary of Defense (Health Affairs)
ATTN: Medical Readiness, Pentagon
Washington DC 20310-0103

Deputy Director for Medical Readiness
The Joint Staff, J-4
ATTN: J4-MRD
4000 Joint Staff, Pentagon
Washington DC 20318-4000

HQDA, Assistant Secretary of the Army
(Research, Development and Acquisition)
ATTN: SARD-TM, Pentagon
Washgton DC 20316-0103

HQDA, Office of The Surgeon General
ATTN: DASG-RDZ
Room 3E368, Pentagon
Washington DC 20310-2300

HQDA, Office of The Surgeon General
ATTN: DASG-ZA
5109 Leesburg Pike
Falls Church VA 22041-3258

HQDA, Office of The Surgeon General
ATTN: DASG-DB
5109 Leesburg Pike
Falls Church VA 22041-3258

HQDA, Office of The Surgeon General
ATTN: DASG-MS
5109 Leesburg Pike
Falls Church VA 22041-3258

HQDA, Office of The Surgeon General
ATTN: SGPS-PSP, Preventive Medicine Consultant
5109 Leesburg Pike
Falls Church VA 22041-3258

U.S. Army Medical Research and Materiel Command
ATTN: MCMR-OP
504 Scott Street
Fort Detrick MD 21702-5012

U.S. Army Medical Research and Materiel Command
ATTN: MCMR-PLC
504 Scott Street
Fort Detrick MD 21702-5012

U.S. Army Medical Research and Materiel Command
ATTN: MCMR-PLE
504 Scott Street
Fort Detrick MD 21702-5012

Uniformed Services University of the Health Sciences
ATTN: Dean, School of Medicine
4301 Jones Bridge Road
Bethesda MD 20814-4799

Uniformed Services University of the Health Sciences
ATTN: Chair, Department of Preventive Medicine
4301 Jones Bridge Road
Bethesda MD 20814-4799

Commandant
Army Medical Department Center and School
ATTN: HSMC-FR, Bldg. 2840
Fort Sam Houston TX 78236-6100

Commandant
Army Medical Department Center and School
ATTN: Chief Librarian, Stimson Library
Bldg. 2840, Room 106
Fort Sam Houston TX 78236-6100

Commandant
Army Medical Department Center and School
ATTN: Director of Combat Development
Fort Sam Houston TX 78236-6100

Commander
U.S. Army Aeromedical Research Laboratory
ATTN: MCMR-UAX-SI
Fort Rucker AL 36362-5292

Commander
U.S. Army Medical Research Institute
of Chemical Defense
ATTN: MCMR-UVZ
Aberdeen Proving Ground MD 21010-5425

Commander
U.S. Army Medical Materiel Development Activity
ATTN: MCMR-UMZ
504 Scott Street
Fort Detrick MD 21702-5009

Commander
U.S. Army Institute of Surgical Research
ATTN: MCMR-USZ
3400 Rayley E. Chambers Avenue
Fort Sam Houston TX 78234-5012

Commander
U.S. Army Medical Research Institute
of Infectious Diseases
ATTN: MCMR-UIZ-A
504 Scott Street
Fort Detrick MD 21702-5011

Director
Walter Reed Army Institute of Research
ATTN: MCMR-UWZ-C (Director for Research Management)
Washington DC 20307-5100

Commander
U.S. Army Soldier Systems Commander
ATTN: AMSSC-CG
Natick MA 01760-5000

Commander
U.S. Army Natick Research, Development
and Engineering Center
ATTN: SSCNC-Z
Natick MA 01760-5000

Commander
U.S. Army Natick Research, Development
and Engineering Center
ATTN: SSCNC-S-IMI
Natick MA 01760-5040

Commander
U.S. Army Natick Research, Development
and Engineering Center
ATTN: SSCNC-TM (U.S Marine Corps Representative)
Natick MA 01760-5004

Director
U.S. Army Research Institute for Behavioral Sciences
5001 Eisenhower Avenue
Alexandria VA 22333-5600

Commander
U.S. Army Training and Doctrine Command
ATTN: ATMD (Office of the Surgeon)
Fort Monroe VA 23651-5000

Commander
U.S. Army Center for Health Promotion
and Preventive Medicine
Aberdeen Proving Ground MD 21010-5422

Director, Biological Sciences Division
Office of Naval Research -Code 141
800 N. Quincy Street
Arlington VA 22217

Commanding Officer
Naval Medical Research & Development Command
NNMC/Bldg 1
Bethesda MD 20889-5044

Commanding Officer
U.S. Navy Clothing & Textile Research Facility
ATTN: NCTRF-01, Bldg 86
Natick MA 01760-5053

Commanding Officer
Naval Environmental Health Center
2510 Walmer Avenue
Norfolk VA 23513-2617

Commanding Officer
Naval Aerospace Medical Institute (Code 32)
Naval Air Station
Pensacola FL 32508-5600

Commanding Officer
Naval Health Research Center
Bethesda MD 20889

Commanding Officer
Naval Health Research Center
P.O. Box 85122
San Diego CA 92138-9174

Commander
Armstrong Medical Research Laboratory
Wright-Patterson Air Force Base OH 45433

U.S. Air Force Aeromedical Library
Document Services Section
2511 Kennedy Circle
Brooks Air Force Base TX 78235-5122

Commander
U.S. Air Force School of Aerospace Medicine
Brooks Air Force Base TX 78235-5000

Director
Human Research and Engineering Directorate
U.S. Army Research Laboratory
Aberdeen Proving Ground MD 21005-5001

U.S. Army Biomedical R&D Representative
Science and Technology Center, Far East
ATTN: AMC-S&T, FE. Unit 45015
APO 96343-5015

National Defence Headquarters
ATTN: Research and Development Branch
Human Performance Directorate
305 Rideau Street
Ottawa Ontario CANADA K1A OK2

Department of National Defence
Defence Research Establishment Ottawa
ATTN: Head, Physiology Group
Environmental Sciences Section
Protective Sciences Division
Ottawa Ontario CANADA M3M 3B9

Director
Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue W.
P.O. Box 200
Downsview Ontario CANADA M3M 3B9

Defence and Civil Institute of Environmental Medicine
ATTN: Head, Environmental Physiology Section
1133 Sheppard Avenue W.
P.O. Box 200
Downsview Ontario CANADA M3M 3B9